

Measurement-based characterization of the load of a Mobile IPv6's Home Agent

Mihai Cristian³, Albert Cabellos-Apariciol¹, Rares Cosma³, Jordi Domingo-Pascual¹, Pedro Vale Pinheiro², Fernando Boavida²

1. Universitat Politècnica de Catalunya
Departament d'Arquitectura de Computadors
.08034 Barcelona, Spain

2. Universidade de Coimbra Departamento de Engenharia Informática
.3030-290 Coimbra, Portugal

3. Technical University of Cluj Napoca, Dept.of Communications.
26-28 George Baritiu St.
Cluj-Napoca, Romania(studying at UPC)

Abstract—Wireless technologies are rapidly evolving and the users are demanding the possibility of changing its point of attachment to the Internet (i.e Access Router) without breaking the IP communications. This can be achieved by using Mobile IPv6. However mobile clients must forwards its data packets addressed towards its home network through a special entity, the Home Agent (HA). This HA is a key point when considering the performance of Mobile IPv6-based networks. This paper presents the firsts steps towards characterizing the load of a HA. This may be useful both for researchers that aim to propose novel architectures that improve the performance of the HAs and for ISPs willing to deploy Mobile IPv6. To achieve our goals first we analyze the internal traffic of a medium-size department. Then we review existing models and evaluate its applicability to this particular scenario. Our results show that the estimated load of a HA serving a medium-size department (around 1500 hosts) is high with a maximum throughput of 262Mbps. Additionally we show that existing models of Wireless LAN networks can be applied to this scenario.

Index Terms—Mobile IPv6, Home Agent, Characterization, Modeling

I. INTRODUCTION

Wireless technologies have rapidly evolved in recent years. IEEE 802.11 is one of the most used wireless technologies and it provides up to 54Mbps of bandwidth in an easy and affordable way. In the current Internet status a user can be connected through a wireless link but he cannot move (i.e. change its access router) without breaking the IP communications. That's why IETF designed Mobile IP which provides mobility to the Internet. With "mobility", a user can move and change his point of attachment to the Internet without losing his network connections.

In Mobile IP a Mobile Node (MN) has two IP addresses. The first one identifies the MN's identity (Home Address, HoA) while the second one identifies the MN's current location (Care-of Address, CoA). The MN will always be reachable through its HoA while it will change its CoA according to its movements. A special entity called Home

Agent (HA) placed at the MN's home network will maintain bindings between the MN's HoA and CoA addresses.

The main limitation of Mobile IP is that communications between the MN and its peers are routed through the HA. This means that a HA may be responsible of multiple MNs on a Home Link. The failure of a single HA may then result in the loss of connectivity of numerous MNs. Thus, HAs represent the possibility of a single point of failure in Mobile IP-based networks. Moreover MN's communications through the HA may also lead to either the HA or the Home Link becoming the bottleneck of the system. In addition, the HA's operation such as security check, packet interception and tunneling might not be as optimized in the HA's software as plain packet forwarding.

Mobile IP comes into two flavors, Mobile IPv4 [13] and Mobile IPv6 [14]. Mobile IPv6 outperforms Mobile IPv4 in many aspects. For instance Mobile IPv6's clients can communicate directly with its peers. This means that these communications are not forwarded through the HA. This reduces the communication's delay and the load at the HA. Unfortunately communications to/from the Home Network must be forwarded through the HA. This paper focuses on a Mobile IPv6's HA.

The research community has focused on solving these issues proposing novel architectures that improve both the performance and the reliability of the HAs[15,16,17,18]. Although they are very effective, Mobile IPv6 has not been deployed yet. This means that the load of a Mobile IPv6's HA is unknown. Hence the proposed architectures might have been evaluated with an unrealistic load. On the other hand ISP's willing to deploy Mobile IPv6 need an estimation of the expected load that the deployed Home Agents will have. That is why we believe that modeling the load of a Mobile IPv6 HA is important for the research community and for the industry.

This paper presents the firsts steps towards characterizing the load of a Mobile IPv6 Home Agent. First we have analyzed the internal traffic of a medium-size department of the UPC. We have assumed that all the hosts inside the department are Mobile IPv6 nodes that are away of its Home Network. Thus all the internal traffic must be processed by an

hypothetical Mobile IPv6's HA. Second we have reviewed existing models of traffic that can be applied to this particular scenario. Specifically we have focused on the models that characterize the load of Wireless LAN networks [1,2]. We have evaluated its goodness-of-fit for our particular case. Finally we have analyzed at which granularity these models can be applied.

Our results show that high processing power is needed when deploying a Home Agent in a medium size scenario. Also we show that existing models for flow-level variables such as flow size and flow inter-arrival times apply to our data on the aggregated traffic level, even if these models have been estimated under different circumstances. Finally we find that these models would also apply to the subnetwork traffic if the number of clients within the subnets is high enough.

The rest of the paper is organized as follows, Section II presents an overall look of the datasets used in terms of internal vs external traffic, transport and application protocols, Section III contains an empirical characterization of the load of our Home Agent, Section IV reviews other models that fit with our estimates for the aggregate traffic, Section V looks to the traffic generated per sub-net and Section VI contains our conclusions.

II. DATA ACQUISITION AND TRAFFIC BREAKDOWN

In this section we present an overlook of the datasets used for this study and focus on the structure of the traffic that is relevant for the goal of this paper. We describe how the traffic was divided in terms of Internal and External traffic and examine the main components of the traces on the transport and application layer.

A. Internal versus External

The data analyzed comprises NetFlow records from a department router at UPC Barcelona, for six days of traffic, from March 9 to March 14, 2007 both inside the department and to/from the Internet. For the 144 hours of traffic monitored we found a total of 903.7 Gigabytes. Figure 1 shows that 95% of all flows, octets and packets belong to traffic to and from the Internet and only 5% represents the traffic between hosts inside the department.

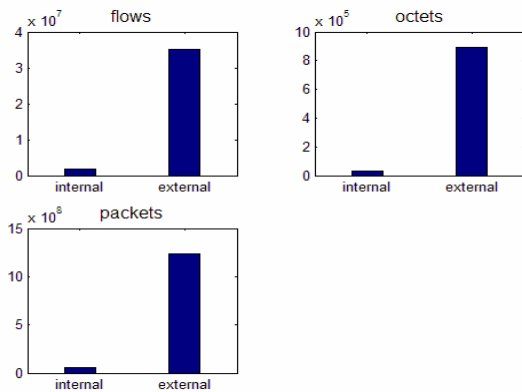


FIGURE 1: INTERNAL VS EXTERNAL TRAFFIC

We base our study on the assumption that all clients inside the department are Mobile IPv6 clients. Following this assumption the traffic between clients inside the department (labeled internal) should be routed by the Home Agent. Traffic to and from the Internet (labeled external) is delivered to its destination directly using IPv6 extension headers and the Return Routability procedure [11]. As our aim is to characterize the load of the Home Agent we disregard the external traffic.

It is worth noting here that Mobile IPv6 clients can also communicate directly with its peers, even if they are at its Home Network. However these communications must be secure and this is achieved by routing them through the Home Agent. Additionally this route does not affect the performance of the communications since the Home Agent and the peers of the Home Network are very close.

Internal traffic is the traffic between Mobile IPv6 clients and their Home Network (the servers in the department). The number of hosts (Mobile IPv6 clients) inside the department sending and receiving packets is 1547. They account for a total of 32.95 Gigabytes of traffic on 1773751 flows. We provide an analysis of this traffic the next sections. These hosts are divided into 7 different subnets each with a number of hosts ranging from 140 to 220 (subnets 1 to 7). The addresses of the hosts have been anonymized.

B. Traffic Breakdown

In order to efficiently characterize the load of the Home Agent we must first understand the makeup of the traffic that we are analyzing. To do this we chose a similar approach to the one presented in [3, 8], that is examining the traffic on the transport and application layer.

In Table 1 we break down the traffic by transport protocol in terms of flows, octets and packets.

As expected most of the octets are sent using the TCP protocol. The percent of the ICMP traffic increases during the second and third day, Saturday and Sunday, but this is only because the total traffic on these days is less than the average (less staff is present from the department during the weekend). The TCP and UDP traffic remains fairly constant during all the days. We find that the bulk of the traffic is sent using the TCP protocol for reasons explained below.

Protocol	flows	octets	Packets
17(udp)	28.75%	0.81%	3.73%
6(tcp)	67.87%	99.13%	95.84%
1 icmp)	3.37%	0.05%	0.42%

TABLE 1: FRACTION OF FLOWS, OCTETS AND PACKETS FOR DIFFERENT TRANSPORT PROTOCOLS

Next we take a look to the application layer. To categorize the traffic inside the department we grouped the applications into several high-level categories.

Table 2 shows the main applications found. The applications have been identified using the flow destination port. Only the most important applications are shown, the ones that account for most of our flows (minor applications have been ignored). The main applications found are consistent with the deployed services in the department.

category	protocols
Bulk	ftp,https,tftp,rtip
Email	smtp,imap,pop,brutus,pop3s
interactive	telnet,ssh
Name	Dns,netbios-ns
net-manage	dhcp,ntp,epmap,snmp,timed
Web	http,https
Windows	netbios-ssn,netbios-dgm
authentication	ident
Printing	lpp
streaming	hp-pdl-datastr

TABLE 2: APPLICATIONS AND THEIR RESPECTIVE PROTOCOLS

Figure 2 shows the application usage for the aggregated traffic (all seven subnets). The percentage ratio changes depending on the week day, as explained earlier. We can clearly see that the most used applications both in terms of flows and packets are email and interactive (especially SSH in our case). In terms of flows email applications represent roughly 50% and SSH only 5%, whereas in terms of octets SSH accounts for almost 60% percent of the traffic, which indicates that it was used for large file transfers in the period monitored. Other applications show a normal ratio between the flows and octets percentage.

The figure also reflects the findings in Table 1. Most of the application layer protocols found use the TCP protocol (web, email, SSH) which explains the 99% octets transmitted using TCP, whereas in terms of flows only 67% percent used TCP, mainly because of the name and net management applications which use the UDP protocol.

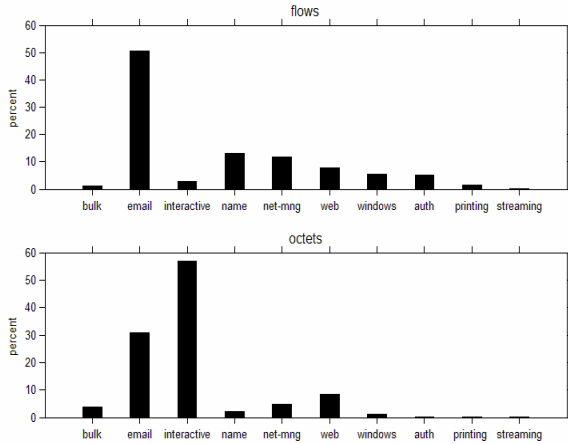


FIGURE 2: APPLICATION USAGE IN TERMS OF FLOWS AND OCTETS

Finally it is worth noting here that [3] shows a similar analysis regarding internal enterprise traffic and finds the same pattern of large numbers of bytes traversing the network on a small number of flows.

III. EMPIRICAL ANALYSIS OF THE LOAD OF THE HOME AGENT

In this section we aim to provide an empirical characterization of the traffic that a Home Agent in a medium size department would have to process. The department contains a number of 1547 hosts (assumed as IPv6 clients), which account for a total of 32.95 Gigabytes of traffic during the one week interval. In order to characterize this load we

have computed a series of parameters such as number of flows per second, flow size, flow inter-arrival times, number of bytes per second and number of packets per second.

We start by looking at the flow arrival process to see if we can observe any patterns that could be helpful in our characterization. Figure 3 plots the time series of the flow arrivals for the entire network using one hour bins. The plot shows sharp increases in the number of flow arrivals in the morning with peaks at 28000 flows per hour during weekdays and 9000 flows per hour during the weekend.

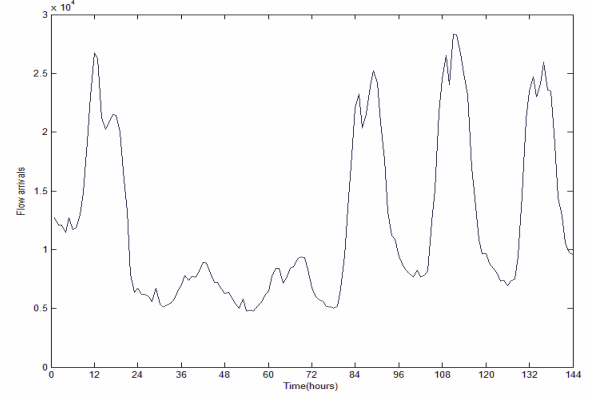


FIGURE 3: FLOW ARRIVALS PATTERN

Another important flow-level variable for the traffic load is the number of flows per second. This parameter represents the number of active flows that, for each second, a HA has to handle. This allows us to see what kind of loads our Home Agent should expect to process when deployed in a similar environment to that of the department. Figure 4 plots the CDF of the average flows per second for each of the six days of traffic. The plots show a clear distinction between the weekdays and the weekend as observed before. For the weekdays we find a mean value of 4.67 flows per second and a maximum of 19.7 flows per second, while during the weekend the mean drops to 1.9 flows per second. The large difference between the weekday and weekend plots also indicates that modeling this variable using a single parametric distribution is rather difficult.

Throughput parameters have also been computed for this scenario for all the days using one minute intervals for averaging. Table 3 summarizes our findings regarding these particular parameters. For each day we compute the mean and maximum throughput in bits and packets per second (bps and pps) and the mean and maximum active flows per second (fps). The mean and maximum values for the throughput are shown in Megabits per second.

day	Mean bps	Max bps	Mean pps	Max pps	Mean fps	Max fps
1	1.39	102.95	217.3	12.66K	4.73	20.8
2	0.03	3.28	18.2	0.29K	1.9	18.41
3	0.04	4.35	22.36	7.54K	1.89	14.98
4	0.68	61.58	123.69	7.33K	4	35.75
5	0.63	262.24	124.26	24.66K	4.67	19.7
6	0.9	123.48	191.58	14.61k	4.27	12.46

TABLE 3: THROUGHPUT VALUES FOR THE AGGREGATED TRAFFIC

We can see that during the six day monitoring period our hypothetical HA would have to process mean values of up to 1.39Mbps and 217.63 pps. Maximum values reach 262.24Mbps and 24.66 Kpps on day 5. We can also distinguish a pattern that holds for all the mean values in the table. Both bps and pps mean values start from a low value in the weekend (days 2 and 3) and increase gradually, peaking on the last day of the week (day 1).

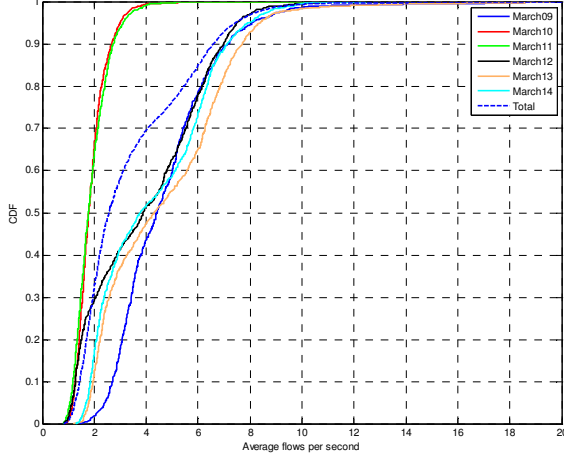


FIGURE 4: DISTRIBUTION OF AVERAGE FLOWS PER SECOND

The next component we look at is the flow inter-arrival time distribution. This parameter is essential for estimating a Kendall queuing model [12] for our deployed HA and was computed in milliseconds. Figure 5 plots the distribution for the six days monitored. We can see that there is still a difference between the weekdays and weekends but it is much smaller than in the case of flows per second or throughput. Mean values for the flow inter-arrival time are 279 ms for the weekdays and 618 ms for the weekend, which is consistent with the plot in Figure 3 where we see a smaller number of flows arriving during the weekend.

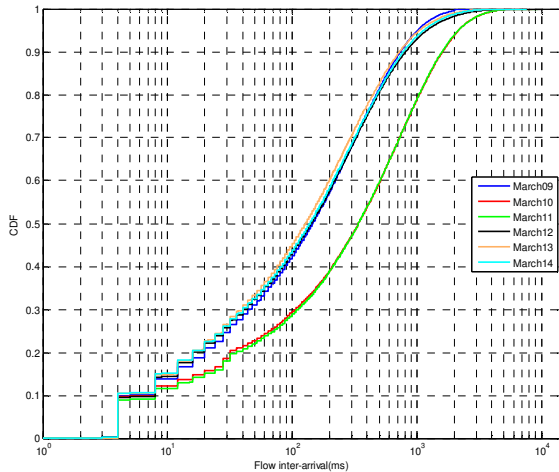


FIGURE 6: CDF PLOT OF FLOW INTER-ARRIVAL TIMES

In order to provide a good characterization of the load of our hypothetical home agent it is not enough to look at the

flow arrival process but also at the flow sizes to find out how much each flow carries in terms of bytes.

Average flow sizes have been computed for one minute intervals, for everyday of traffic, by dividing the total number of octets to the total number of flows inside the bin. The average flow size distribution is plotted in Figure 7. Flow sizes show mean values of 10.4 Kilobytes per flow during the weekdays and 2.47 Kilobytes per flow in the weekend.

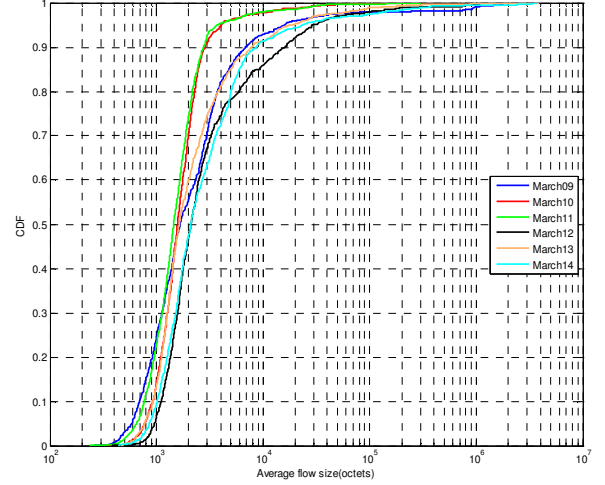


FIGURE 7: AVERAGE FLOW SIZE DISTRIBUTION FOR ALL DAYS OF TRAFFIC

Summarizing, this section attempts to provide an empirical characterization of the load of a HA. With a total of roughly 1500 Mobile IPv6 nodes, a HA would have to process a mean value of 3.57 active flows per second and should handle mean throughput values of 0.67Mbps and 120 pps. Maximum throughput values of 262.24Mbps indicate that high processing power is needed to deploy as HA in a similar scenario. Also we have spotted patterns in the the flow arrival process, flow inter-arrival times and the average flow-sizes for hour traffic which indicate that these variables can be modeled using parametric distributions.

IV. MODELING THE DATA

To the best of our knowledge no studies exist aiming to characterize the load of a Mobile IPv6 Home Agent. Nevertheless other researchers have characterized similar loads. In this section we aim to evaluate if these existing models fit on our empirical data.

A similar analysis to ours is found in [1]. However the authors attempt to model the traffic for a campus WirelessLAN, not for a Home Agent as is our case. Their approach is based on two levels of modeling: the session and the flow level. Flow arrivals in [1] are considered as a cluster process triggered by session arrivals which is not the case for our study. We focus our analysis entirely on the flow level. Additionally the load characterized both in [1] as well as our analyzed data are highly dependent on the applications deployed. Application usage differs when looking at a Wireless LAN or a HA load which could lead to different load characteristics. We will try to see if the models proposed fit with our empirical data.

We find that our flow arrivals pattern (Figure 3) is very similar to the session arrivals pattern depicted in [1] which leads us to believe that models for other flow variables might apply in our scenario as well. Variables on the flow level have been modeled for WLAN traffic using the following distributions: flow-size using a Pareto distribution and flow inter-arrival time using the log-normal distribution.

We will evaluate the fits of the proposed distributions to see if they match our findings. Figure 8 plots the flow inter-arrival empirical distribution against the proposed LogNormal distribution with the following parameters: mean=365.72, $\mu(\log \text{ location})=5.77$ and $\sigma(\log \text{ scale})=0.49$, with its 95% confidence bounds. The empirical distribution remains within the confidence bounds of the LogNormal distribution, but in the tail section, which contains mostly extreme values, it varies from lognormality.

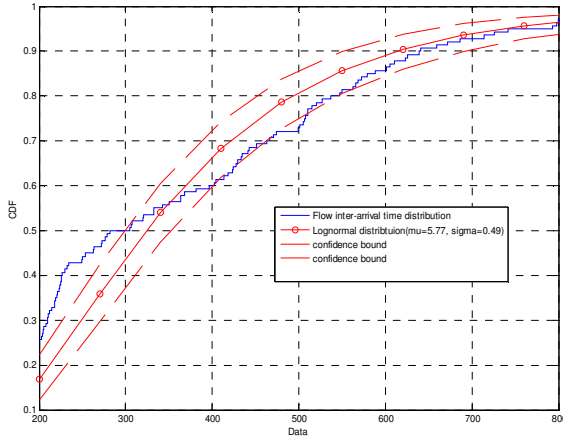


FIGURE 8: EMPIRICAL DISTRIBUTION OF FLOW INTER-ARRIVAL TIMES PLOTTED AGAINST LOGNORMAL DISTRIBUTION

When attempting to fit the proposed distribution in [1] for the flow size empirical data we come across the same issues. Figure 9 plots the empirical distribution of the flow sizes with its confidence bounds against the proposed Pareto distribution with the following parameters: mean=3099.9, $k(\text{shape})=0.63$, $\sigma(\text{scale})=1234.18$ and $\theta=635$. Again we can see that the proposed model provides a good fit for our empirical data, with the same problem in the tail of the curve where the empirical distribution has a higher skew.

We also found that the generalized extreme value distribution provides a slightly better fit especially for the mentioned tail section because this section contains mostly extreme values of the plotted data.

Our findings indicate that flow variables for the load of a HA can be modeled using statistical distributions. Existing models found in [1] provide a good fit for our empirical data and can be used to model the load of a hypothetical HA in a scenario with parameters similar to ours (number of hosts and deployed services).

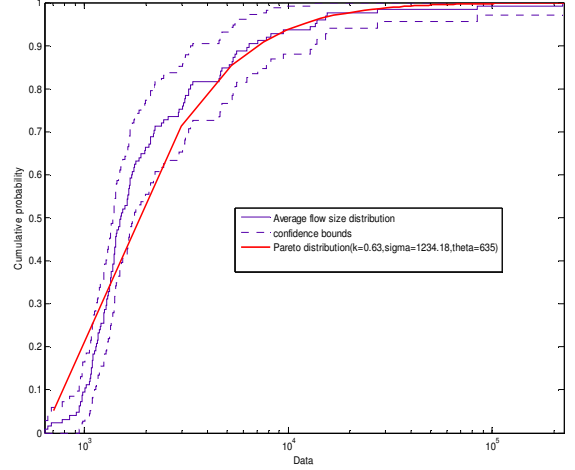


FIGURE 9: EMPIRICAL DISTRIBUTION OF FLOW SIZES PLOTTED AGAINST GENERALIZED PARETO DISTRIBUTION

V. SUB-NETWORK TRAFFIC

In this section we try to find patterns in the per-subnet traffic and fit them to the above mentioned models. This could be useful for adapting the models to suit a scenario with any given number of subnetworks. The same parameters presented in Section III have been computed for each of the seven subnetworks in the department using one minute intervals as stated before.

Figure 10 plots the average flow size distribution for each of the seven subnetworks using dotted lines and the average flow size for the aggregated traffic using a solid line. For this parameter we can see that subnet 1 and 4 follow the Pareto distribution of the aggregate traffic. Other subnets can be grouped together according to their distribution, for instance subnets 5 and 3 have similar flow size distributions, but subnet 6 and subnets 7 and 2 do not resemble the model of the aggregated traffic.

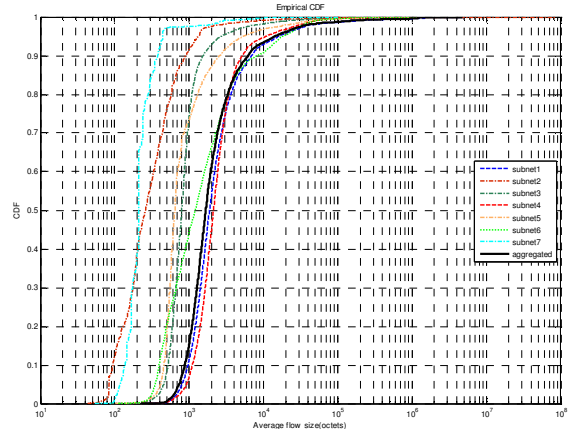


FIGURE 10: AVERAGE FLOW SIZE PER SUBNETWORK

In terms of flow inter-arrival times, plotted in Figure 11, we see a higher resemblance between the distribution of the subnets and the LogNormal distribution of the aggregated traffic. As before subnets 1 and 4 show the closest match whereas subnet 2 shows a completely different distribution.

The large difference between distributions for sub-networks can be explained by the fact that they have different application usage patterns. A similar analysis to the one in Section II revealed that subnets 1 and 4 present the highest resemblance to the overall application usage (Figure 2), whereas other subnetworks show a completely different pattern. Also subnet 1 and subnet 4 have the highest number of clients.

This leads us to believe that the flow-level traffic models presented for the aggregated traffic should work for subnetworks with a higher number of clients. A higher number of clients leads to a smoother curve for the evaluated parameters and an application usage pattern that complies with the one presented for the aggregated traffic.

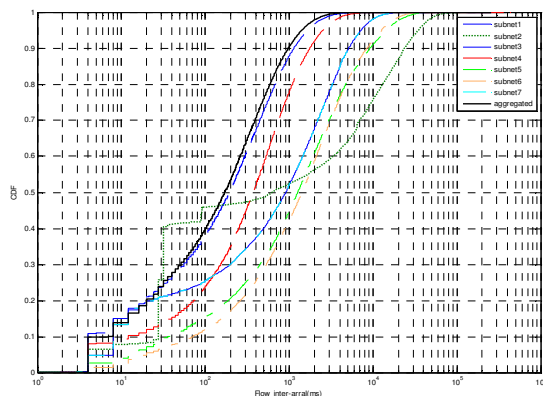


FIGURE 10: DISTRIBUTION OF FLOW INTER-ARRIVALS PER SUBNET

VI. CONCLUSIONS AND FUTURE WORK

In this paper we presented the first steps towards characterizing the load of a Mobile IPv6's HA. In order to do this we have measured the internal traffic of a medium-size department and assumed that all the hosts are Mobile IPv6 clients away from home.

In the first part of the paper we evaluate the load that a Mobile IPv6 Home Agent would have to process. Our results show that in a medium size network comprising 1500 clients a HA is looking at mean values of 3.57 active flows per second and maximum throughput values of 262.24 Mbps and 24.6 Kpps. This indicates that high processing power is needed for a HA deployed in a similar scenario.

The second part of our study looks to fit existing traffic models to our empirical data. We found that variables such as flow-size, inter-arrival times and flow arrivals fit currently deployed models for the aggregated traffic even if these models have been developed for characterizing Wireless LAN traffic and not the load of a HA. The following variables have been modeled: average flow sizes using a Pareto distribution and flow inter-arrival times using a Lognormal distribution.

In the final part we look at the per subnet traffic for similar patterns. We find that the same distributions can be used to model the traffic at the subnet level for subnetworks with a high number of clients. Further work is needed in this direction to match subnet traffic patterns with aggregated traffic patterns.

REFERENCES

- [1] F.Hernandez-Campos, M.Karaliopolus, M.Papadopouli, H.Shen. "Spatio-Temporal Modelling of Traffic Workload in a Campus WLAN", WICON 2006
- [2] M. Karaliopolus, M. Papadopouli, E.Raftopolous, H.Shen. "On scalable measurement-driven modelling of traffic demand in large WLANs", LANMAN 2007
- [3] R.Pang, M.Allman, M.Bennet, "A first look at modern enterprise traffic", IMC 2005
- [4] A.Sinha, K.Mitchell, D.Methi. "Flow level upstream traffic behaviour in broadband access networks: DSL versus Broadband fixed wireless", IPOM 2003
- [5] C.E. Perkins. "Mobile IP", IEEE Communications Magazine, May 1997
- [6] Mark Fulmers flow-tools
- [7] Z.Sun, D.He, L.Liang. "Internet QoS and traffic modelling", IEE 2004
- [8] M.Ploumidis, M.Papadopouli, T.Karagiannis. "Multi-Level application-based traffic characterization in a large scale wireless network", WOWMOM 2007
- [9] K.Thompson, G.J.Miller, R.Wilder. "Wide-Area Traffic Patterns and Characteristics", 1997
- [10] C.E.Perkins, D.B. Johnson. "Mobility Support in IPv6", MOBICOM 1996
- [11] K.J.Lee, J.Park, H.Kim. "Route optimization for mobile nodes in a mobile network based on prefix delegation", VTC 2003
- [12] N.Vandaele, T.Van Woensel, A.Verbuggen. "A queueing based traffic model", March 2000
- [13] D.Johnson et al. "Mobility Support in IPv6", RFC 3775, 2004
- [14] C. Perkins: IP Mobility Support for IPv4 RFC 3344 (2002)
- [15] F. Heissenhuber et al. "Home Agent Redundancy and Load Balancing in Mobile IPv6" I.Conf Broadband Communications 1999
- [16] H. Deng et al "Load Balance for Distributed Home Agents in Mobile IPv6", IEEE PIMRC 2003
- [17] R. Wakikawa et al "Virtual mobility control domain for enhancements of mobility protocols" INFOCOM 2005
- [18] Albert Cabellos-Aparicio et al., "A flexible and Distributed Home Agent Architecture for Mobile IPv6-based Networks" IFIP Networking 2007